DOCUMENT RESUME

ED 338 670 TM 017 475

AUTHOR Coffman, William E.
TITLE Testing in the Schools: A Historical Perspective.
INSTITUTION Center for Research on Evaluation, Standards, and Student Testing, Los Angeles, CA.
SPONS AGENCY National Inst. of Education (ED), Washington, DC.
REPORT NO CSE-R-253
PUB DATE 85
NOTE 25p.; Paper presented at a Conference on "Paths to Excellence: Testing and Technology" (Los Angeles, CA, July 14-15, 1983).
PUB TYPE Information Analyses (070) -- Viewpoints (Opinion/Position Papers, Essays, etc.) (120) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Classroom Research; Educational Assessment; *Educational History; Educational Policy; Educational Research; Educational Trends; Elementary Secondary Education; Literature Reviews; *Public Policy; Researchers; *Test Use; *Theory Practice Relationship; Trend Analysis
IDENTIFIERS Personal Experiences

ABSTRACT

The experience of the speaker is applied to the changes he has seen in the field of testing over his career. Many of the concepts, issues, and controversies engaging the educational research community today had already been identified at the beginning of the speaker's career in the 1930s. A review of the literature of the past 50 years reveals one common thread: concern on the part of measurement specialists that teachers seem not to be taking seriously the admonitions of researchers and measurement specialists regarding ways of using tests in the classroom. Other common threads are seen in the study of the relationship between test theory and practice, and the relationship between testing and public educational policy. A survey of the literature related to educational testing, as filtered through the observations of one person over 50 years, suggests that the responses to questions do not have much meaning unless they are placed in context. The ways in which tests are really being used is the essential point. A 37-item list of references is included. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

C·R·E·S·S·T

Center for Research
on Evaluation, Standards,
and Student Testing

ED388670

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

William E. Coffman
E.F. Linquist Professor Emeritus
University of Iowa

TM017475

BEST COPY AVAILABLE

Testing in the Schools: A Historical Perspective

William E. Coffman
E.F. Linquist Professor Emeritus
University of Iowa

CSE Report No. 253
1985

Center for the Study of Evaluation
Graduate School of Education
University of California, Los Angeles

The project presented or reported herein was in part performed pursuant to a grant from the National Institute of Education, Department of Education. However, the opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education and no official endorsement by the National Institute of Education should be inferred.

TABLE OF CONTENTS

	Page
Introduction	1
Themes, Developments, and Cycles	4
Teacher Education in Testing	6
Teachers and Researchers	7
Different Philosophical Positions	10
Testing and Public Policy	13
Conclusions	15
References	18

INTRODUCTION

> Teachers are important people. They are the people directly responsible for the education of the children and youth of our country. The curriculum of the school is largely what they make it. The professor of education, the school administrator, or the curriculum director may have a large part in determining the content of printed courses of study. They may be responsible for much of the talking and writing in the field of education. But what goes on in the school depends on the teacher in the classroom--on the way he accepts and implements the ideas of the experts or adds his own creative touch based on his unique experience with a particular group of pupils. The teacher, then, is a key person in any program of curriculum development (Coffman, 1951, p. 305).

I wrote these words a long time ago and in a context different from that of today's conference.* But I believe that with a little modification they can be made relevant to the topic of testing in the schools today. Teachers are indeed important people, not only in determining the actual curriculum but also in determining how tests are used in relation to teaching and learning. The legislator, in Washington or the state capitol, may pass laws that mandate specific testing programs; school administrators, in the Department of Education of the nation or state, or of the local school system, may publish edicts or require periodic reports; experts in educational and psychological measurement may argue issues, collect data and publish interpetation, and admonish teachers to do this or that; but, at least in most educational settings, what actually happens is determined by teachers as they interact with pupils in classrooms. One

---

*This paper was first presented at a conference "Paths to Excellence: Testing and Technology" hosted by the UCLA Center for the Study of Evaluation (CSE), July 14-15, 1983.

might, therefore, with good reason, ask why it is that so little hard data are available on what actually does happen. And if one wants to make sense of the limited data that are in hand, how must they be organized and interpreted?

I found myself searching my own professional experience for answers to these questions, and then checkng my impressions by referring to more than a half century of published literature. The year I made the decision to enter the field of education, 1931, was the first year of publication of the Review of Educational Research; and two years later the February issue provided the first review on the topic "Educational Tests and Their Uses", a review that cited 467 references (Wood, 1933). The Education Index first appeared in 1929, and the first bound volume in the University of Iowa library (January 1929-June 1932) contains entries under the headings "Examinations" and "Tests and Scales" that reflect interest in and concern with issues still of relevance today: "Examinations as an aid to learning" (Jersild, 1929), "Examinations seventy-five years ago and today" (Fish, 1930), "Conflicting philosophies concerning educational measurement" (Brown, 1931), "History of the measurement movement" (Malin, 1930), and "Participation in testing programs by the classroom teacher" (Macken, 1929). The heading "Evaluation" first appeared in the next bound volume (July 1932-June 1935), but there was only one entry. Entries increased rapidly during the late 1930's and through the 1940's as concerns broadened to educational outcomes other than recall of information.

The Review of Educational Research carried reviews concerned with testing in the schools at approximately three-year intervals until a more focused and less comprehensive format was adopted during the 1970's. The

Education Index marked the growing complexity of the field by expanding the variety of headings, as did the Encyclopedia of Educational Research, beginning with the first edition in 1941. From time to time, the National Society for the Study of Education focused on research and testing in one or another of its yearbooks. And more recently, the annual Review of Research in Education and the ERIC publications have helped us keep on top of a proliferating literature.

The span of my own professional career covers the period since these systematic reviews first appeared in the literature. The first third of the period since then (1931-1949), I was a classroom teacher and administrator in public schools. Since 1949, I have worked as a specialist in measurement and evaluation. The literature, then, serves to confirm, deny, or expand my own recollections.

This is not to say that measurement first became a topic of concern to educators in the 1930's. I note, for example, that the Twenty-First Annual Conference of Educational Measurement was held at the University of Indiana in 1934, and that Scates was looking back over a period of 50 years as early as 1947 (Scates, 1947). But conferences are often more opportunities for the sharing of impressions than for the reporting of solid evidence, and histories can focus on the highlighting of deficiencies and admonitions for sounder procedures in the future than on the documentation of accomplishments. It was certainly very soon after the accumulated literature began to be systematically reviewed that the scientific movement in education came of age (NSSE, 1935; 1938), and the decade of the 1930's was particularly productive in new insights and challenges. As one of the

leaders in the organization of the educational research profession noted at the time,

> Each generation seems to discover for itself teleological and methodological concepts which it brands as new, or progressive, even though these very ideas may have been formulated and voiced centuries or millenniums earlier. It is difficult to know what is new; most ideas are new only to individuals. It appears, however, that there are strong movements in education today which are actually affecting practice in conventional schools in ways which heretofore was only talked about, or practiced in a few private schools (Scates, 1938, p. 523).

It might be profitable for today's educational researchers, many of whom have brought the conceptual framework and methodological concepts of other academic fields to the study of educational problems, to become acquainted with the educational research literature of the 1930's. The vocabulary may be different, and the total context may be less well-defined than that of today; but the underlying concepts and ideas may often be the same as those that guide today's research.

THEMES, DEVELOPMENTS AND CYCLES

As I have already implied, many of the concepts, issues, and controversies that engage the educational research community today had already been identified early in the 1930's. One can trace these through the literature. In some cases, one finds recurring themes such as a concern with the possibility that standardized tests may have undesirable effects on school curricula. Sometimes there appears to be cyclical movement as a concern shifts from a focus on minimum essentials to a concern with personality development and back again to minimum essentials.

In rare instances, one can detect what appears to be real progress, but the progress is more likely to be in a wider dissemination of insights than in the originality of the insight.

For instance, the beginning of concern for efficiency in education through application of principles from business and industry has been attributed to a paper by Franklin Bobbitt in the 12th Yearbook of the National Society for the Study of Education (1913). In that paper he urged careful specification of what pupils were expected to learn in school, and implied that once objectives were specified, teachers might reasonably be held accountable for seeing that they were achieved. One can see the roots of much of today's concern about minimum essentials in the writing of disciples of Bobbitt over the years. But disciples seldom encompass the full vision of the master, and it is instructive to read what Bobbitt had to say about the importance of considering higher as well as lower level objectives:

> The higher, however, must (also) be scaled. However difficult it may seem to set up quantitative standards in the more intangible field, it must of necessity be done, if once they are introduced into the lower, more objective and more mechanical forms of training. It will work harm to establish definite standards for only a portion of education, leaving the rest to traditional vagueness and uncertainty of aim...But education must take care of all desirable aspects of human personality--training and developing each in due porportion, slighting nothing, neglecting nothing, giving unduly large or unduly small attention to nothing (p. 26).

Bobbit recognized that it wouldn't be easy to quantify the intangible objectives, and the concern he expressed is still with us today. Much of the controversy over educational measurement in the schools since that time

has been concerned with the effect of imbalance in the use of tests, and people are still trying to provide measures of higher level outcomes to redress the balance.

As one prepares to look at testing practices in the schools of the 1980's, it will be profitable to review briefly some of these trends over the years, and to consider their implications for interpeting what we see. Let us begin by considering what we know about teachers' preparation for using tests.

TEACHER EDUCATION IN TESTING

At the time that I completed my undergraduate program in secondary education, my home state of West Virginia required that all applicants for certification as a teacher in the secondary schools had completed a course in tests and measurement. I was enrolled in a college in Ohio, and since Ohio did not have such a requirement, I completed the requirement through individual study. At the time, the fact that such a requirement was not widespread was of little significance to me; but what about now? Apparently, the passing years have not seen much change in the situation. At mid-century, Betts (1950) was taking a dim view of the ability of teachers to interpret standardized test results:

> Such norms (GE) are highly satisfactory to teachers because pupils in general make greater progress during the course of the year than is shown in cross-sectional norms. When standardized testing is done at the beginning of the school year, teachers using the test find a majority of their pupils above the norm at the end of the school year and glow with success. They are unaware that the test they are using probably measures intelligence, not school taught learnings, and that what appears to be greater

> than normal progress, is a mere statistical artifact (p. 218).

In 1959, Mayo reported a study by Noll indicating that 83% of 80 colleges he had surveyed offered a course in measurement, but that only 14% of them required one of all teacher education students. Furthermore, only 10% of the states required a course for certification. Ten years later Stinnet (1969) made no mention of any requirement in educational measurement in his encyclopedia article on teacher certification, nor did Burdin (1982) thirteen years later. It seems obvious that only a minority of teachers have had any intensive training in educational measurement. Is it possible that those who have may exhibit quite different practices from those who have not? Certainly, information regarding the background in educational measurement of respondents would appear to be critical in the interpretation of survey reponses.

To those of us in the measurement profession, the lack of course work in the field in programs of teacher education appears to be a serious omission. The fact that it apparently does not seem so to other educators suggests a need to look more closely. What does such a look reveal?

TEACHERS AND RESEARCHERS

One thread running through the measurement and evaluation literature is a concern, on the part of measurement specialists, that teachers seem not to be taking seriously the admonitions of researchers and measurement specialists regarding ways of using tests in classroom settings. The concern seems seldom to have led to the collection of hard data. One explanation for this phenomenon may be found in an analysis of the problem

by Scates (1943). Scates pointed out that the scientist is interested in truth leading to broad generalizations, while the teacher seeks information of direct practical value; the scientist is interested in elements, whereas the teacher is interested in functioning organisms; the measurement specialist cannot measure continuously, but the teacher needs to and must measure continuously; the scientist measures traits uniform throughout their range, but the teacher measures growth in stages; and the measurement specialist generally measures formal abilities by cross-sectional power tests, but the teacher must be concerned with behavioral dynamics in life situations.

To the extent that Scates's analysis is sound, it is not suprising that there is little systematic study of teachers' testing practices reported in the literature written primarily by researchers and test specialists. They had their own interests, which were different from those of teachers, and they probably weren't even aware that the difference existed.

It is true that over the years the interests of researchers have turned more from concern with simple elements to concern for the dynamics of learning. Still, recent articles tend to confirm the conclusions of Scates:

> Teacher preference, in effect, is for continuous movies, in color with sound, while a test score, or even a profile of scores, is more akin to a black-and-white photograph (Salmon-Cox, 1981).

There is even a tendency to focus on uses of tests in research and guidance rather than as tools in the instructional setting. For example,

> Two functions of tests that deserve particular emphasis at this time are: first, the uses of

> educational tests in the construction and evaluation of educational theories, especially theories that give particular attention to processes or strategies of problem-solving rather than outcomes alone; and second, the uses of tests in the service of individual students through systems of guidance that employ measurement as a means of fostering self-discovery and as a means for encouraging students to develop wisdom in decision-making (Manning, 1970, pp. 20-21).

To some extent, recent interest in qualitative methods have brought the data collection procedures of the researcher closer to the interests of the teacher (Hamilton et al., 1977). But it is unlikely that teachers generally will seek greater expertise in anthropological methods than they have in psychometric methods. It is more likely that if they wish to increase the use of tests in instructional settings, researchers will need to be asking themselves: what is it in our materials and methods that is likely to be useful to teachers whose basic guides to decisions are the moment-by-moment observations so clearly described by Jackson (1968) in Life In Classrooms. And the researcher interested in how teachers use tests will want to collect enough information about the total mix of data, observation as well as formal and informal, testing to understand the place of testing in the mix.

Incidentally, it appears that often the teacher's orientation is different, not only from that of the researcher and test specialist, but also from that of the school administrator and school board member. This idea is well expressed by Gorton (1982, p. 1906):

> Teachers tend to emphasize such aspects as humanistic orientation to instruction and positive relations between teachers and students; administrators, on the other hand, stressed such factors as student achievement on standardized tests and administrative evaluation.

Given that such differences do exist (the research tends to be based on small and often non-representative samples), recent trends toward differentiation of testing in relation to function would probably be welcomed by teachers. Lefever (1950) expressed the possibilities quite clearly almost 25 years ago. He argues (but with no supporting data) that teacher-made tests should be considered essential tools for checking pupil achievement, particularly at the secondary school level; that teachers grow in professional competence as they participate in test construction; that specialists in measurement should be active in in-service education to facilitate sound teacher activity; that general survey testing to evaluate educational programs should never be broken down to the individual class level and might well be conducted using matrix sampling; and that it is essential for teachers to be actively involved in planning the system testing program. To the extent that separation of function of this sort is operating, responses of teachers to survey questions may be expected to differ from those under different circumstances.

DIFFERENT PHILOSOPHICAL POSITIONS

Another issue that has complicated the picture of testing in the schools involves much more than differences between teachers and test specialists, or between teachers and administrators. In fact, there is almost never a simple contrast, for within each of these groups there are likely to be differences about the purposes of education, the nature of human learning, and the nature of evidence, that is, differences in basic philosophy (Coffman, no date; Hughes, 1934; Thelen, 1969; Weiss, 1981).

While the proportions of each group holding a particular position may vary, all positions are likely to be found within each group. Furthermore, the philosophical domain is not a simple one that can be represented by a single dimension, for example, conservative-liberal. In most cases, one needs to look for various dimensions.

There is, for example, the issue of whether the school should be concerned primarily with the transmission of the culture to each new generation or primarily with the development of skills needed for adjusting to a constantly changing culture. There seems little doubt that Bobbitt (1913) was concerned primarily with the former, although his view of the culture to be transmitted was broader than that of many of his followers. Findley and Smith (1950, p. 63) called attention to a contrasting position argued by Brownell (1948). They wrote:

> Brownell offered a criticism of learning implicit in most educational measurement. He insisted that we raise our sights from measures of rate and accuracy of performance to measures of level of process used, from evidence of immediate gains to that of more permanent gains, and from ability to use learning in closely similar situations to transferability to essentially new situations, especially after a significant lapse of time.

More than a decade earlier, Brownell (1937, p. 492) had posed a challenge to test developers that is still challenging them today:

> To meet the proposed criteria, a test must (1) elicit from pupils the desired types of mental process, (2) enable the teacher to observe and analyze the thought processes which lie back of the pupils' answers, (3) encourage the development of desired study habits, (4) lead to improved instructional practice, and (5) foster wholesome relationships between teacher and pupils.

Snow, writing in 1980, sounds the same note, but perhaps the tools for tackling the problem are more appropriate than they were in 1937.

> If one looks only at immediate achievement, ignoring aptitude, and most instructional research still does both of these things, then elaboration of instruction appears beneficial. If one adds general ability to the picture, it turns out that elaboration helps the less able learners but may not be optimal for the most able learners. If one must further choose a particular form of elaboration to give less able students, it appear best to match the form to the learner's relative strengths. However, when retention is considered, all this changes. Unelaborated instruction is best for almost everybody, and particularly for students high in verbal-crystallized ability. And if one had to choose a form of elaboration, it would seem best to mismatch the form with a student's ability profile (p. 56).

Other researchers and test specialists are also showing an interest in the development of tests that can provide data directly applicable to issues in testing and learning (Anderson, 1972; Calfee, 1981; Messick, 1983). In each case, however, the concern is with education designed to develop intellectual skills rather than to transmit information. To teachers who accept the skills objectives, the message in the literature is likely to be significant. To those whose orientation is toward content as the focus of education, the message may have little impact. And what about those holding other positions: that the purpose of education is the cultivation of well-adjusted, happy individuals, or the building of a new social order?

The concern with personality development that characterized the progressive education movement in the 1930's does not seem to be of much concern to researchers and testers today, but there are undoubtedly many with roots in this position who occupy teaching positions today and whose philosophical orientation leads them to the view that tests that focus only on either information or intellectual skills are restrictive. To them, the methods of the clinician are preferable to those of the psychometrician,

and their responses to questions about testing and evaluation will make sense only when the philosophical context is made explicit. They might, however, be surprised to read this quotation from Wood's article in the Review of Educational Research in 1933:

> ...the highest purpose and ultimate aim of the objective testing movement is not to make better college entrance or course-credit examinations, but to help inaugurate a continuous study of individuals throughout the whole educational ladder by means of systematically recorded comparable measures and observations which will make such spasmodic examinations largely unnecessary...The first question that the school should ask and answer at least provisionally several times a year is, "What can Johnny learn, and which of the things he can learn should the school, in the light of all the facts, try to help him learn?" Tests should first of all tell what a pupil should try to learn--not how he may be cajoled, persuaded, or insidiously coerced into the learning item x in the "standard" curriculum for grade n (pp. 7-9).

TESTING AND PUBLIC POLICYy

One factor that may well influence the reactions of teachers to test and evaluation practices, and so be critical to the interpretation of research concerned with the use of tests, is the extent to which policy decisions by public agencies depend on test results. Traditionally, in the United States, policy decisions regarding schooling have rested in the hands of local agencies, and for such decisions, little use has been made of formal testing. In the continuing discussion of ways in which tests might influence teaching practices, there has been recognition of the need to guard against giving too much weight to test results. In fact, as early as the mid-1930's, when Lindquist was establishing the Basic Skills Testing

Program in Iowa, he cautioned that test results, if they were to be useful in guiding teaching and learning, should not be used for the purpose of evaluating teachers or for rating schools (Peterson, 1983). Early studies of teacher practices and attitudes were carried out in this context, and interpetations of results even as late as 1981 may be reflecting to a certain extent the tradition of local control and autonomy. Miller (1963) indicated that in spite of claims to the contrary, there was little likelihood that state or national testing programs would influence very much the practices of good teachers in the secondary schools. Goslin (1967) reported that many teachers look on tests as of peripheral importance. Salmon-Cox (1981) reported that teachers prefer to depend on their own judgment rather than on test results. However, these studies represent another time--or were based on highly specialized samples. The possible effects of recent trends was clearly recognized by Madaus (1981), who wrote:

> U.S. education is now adopting a new relationship between testing and policy, and hence between test results and their use. Testing is now being asked to assume a new role, one in which a test mandated by a policy board (often external to the local school district) becomes the administrative device through which a particular educational policy is implemented. The effects of such testing programs on the balance of power between local districts and the agency mandating the test are a direct function of the rewards or sanctions associated with test use. Both history and the contemporary experience of western European countries reveal that, whenever test results become a key element in important decisions that affect individual life chances (e.g., graduation from high school or grade-to-grade promotion, teacher salary or tenure decisions, school certification, or the allocation of funds), the agency that administers the test assumes a great deal of power over the schooling process. When external tests are used in

> these ways, administrators, teachers, and pupils take the results seriously and modify their behavior and attitudes accordingly. (1981, p. 635)

It would appear, then, that for any clear interpretation of data based on surveys of teacher attitudes and practices with respect to tests and testing, it would be important to assess the extent to which respondents were feeling the effects of the use of tests for implementing policy.

CONCLUSIONS

What, then, does a survey of the literature related to testing in education (when filtered through the collected observations of one person over 50 years) suggest to researchers today seeking insights into how teachers collect and interpret data about pupil achievement? Perhaps the most important conclusion is that one can't make much sense out of responses to questions unless they are placed in an appropriate context. Answers to questions will vary, and the meaning of those answers will depend on a variety of factors affecting the respondent. The interesting findings will be the interactions between questions and these factors, not the first order responses. More specifically, this review suggests that the researcher of the 1980's should consider these things:

1. Studies in the past of teachers' use of tests have been of two kinds. There have been intensive studies of small and non-representative samples that provide a rich framework for interpretation but leave the reader with the feeling that what the researcher found may be true of these teachers in these settings, but not necessarily of other teachers in other settings. There have also been large-scale surveys that break down responses along

easily identified but not necessarily significant categories such as sex, geographical region, level of education, or size of school or community. What is needed is information based on a comprehensive and representative sample that can be broken down along meaningful dimensions.

2. One factor that may well moderate teacher attitudes and practices may be the extent of training in principles of measurement and evaluation. The evidence is that teachers with formal course work in measurement and evaluation at the preservice level are a minority, and that inservice programs vary all the way from extensive and profound to superficial or nonexistent. It will certainly be helpful in making sense of responses to have information about the respondents' background in testing.
3. The literature documents the rather dramatic difference in the views of teachers and researchers regarding what tests should provide in the way of information. Thus, researchers should be on guard against framing survey questions that may be significant to them but not necessarily to teachers--or against framing questions that may be perceived differently by teachers than intended by the researcher. Researchers might even consider researching the question of whether or not the continuous observation described by such researchers as Jackson or Salmon-Cox may be providing teachers with more valid data than that provided by any single test, however comprehensive.

4. Even though teachers and researchers, or teachers and administrators, or teachers and laymen, may differ in general in their attitudes toward testing, there will be, in each situation, philosophical viewpoints that are influencing attitudes and values--and practice. Responses may be different, depending on the philosophy of education of the respondent; and for teachers with the same philosophy of education, responses may differ depending on whether or not that philosophical position is held also by administrators in the system or by officials outside the system who are perceived as holding power over the system. The phenomenal field of the respondent needs to be assessed if responses are to be properly interpreted.
5. Finally, the researcher will need to assess carefully the extent to which the use of tests in the implementation of public policy is having an impact on testing in the schools from which respondents are coming. It is not yet clear whether the increased use of tests for such purposes is a trend that will continue, or whether we are near the peak of a fluctuating cycle. In any case, how the teacher or administrator views the distribution of power may well influence the responses collected by the researcher.

References

Anderson, R.C. (1972). How to construct achievement tests to assess comprehension. Review of Educational Research, 42(2), 145-170.

Betts, G.L. (1950). Better interpretation and use of standardized achievement tests. Education, 71(4), 217-221.

Bobbitt, F. (1913). Some general principles of management applied to the problems of city-school systems. In S.L. Parker (Ed.), The twelfth yearbook of the National Society for the Study of Education (Part I). Chicago: University of Chicago Press.

Brown, M.E. (1931). Conflicting philosophies concerning educational measurement. Kadelpian Review, 10, 175-179.

Brownell, W.A. (1937). Some neglected criteria for evaluating classroom tests. National Elementary Principal, 16, 485-492.

Brownell, W.A. (1948). Criteria of learning in educational research. Journal of Educational Psychology, 39, 170-182.

Burdin, J.L. (1982). Teacher certification. In H.E. Mitzel (Ed.), Encyclopedia of educational research (5th ed.). New York: The Free Press.

Calfee, R. (1981). Cognitive psychology and educational practice. In D.E. Berliner (Ed.), Review of research in education (No. 9). Washington, DC: American Educational Research Association.

Coffman, W.E. (1951). Teacher morale and curriculum development: A statistical analysis of responses to a reaction inventory. Journal of Experimental Education, 19, 305-332.

Coffman, W.E. (1969). Concepts of achievement and proficiency. In P.H. Dubois (Chmn.), Proceedings of the 1969 Invitational Conference on Testing Problems. Princeton, NJ: Educational Testing Service.

Findley, W.G., & Smith, A.B. (1950). Measurement of educational achievement in the schools. Review of Educational Research, 20(1), 63-75.

Fish, L.J. (1931). Examinations seventy-five years ago and today: Comparison and results. Elementary School Journal, 31, 330-332.

Gorton, R.A. (1982). Teacher job satisfaction. In H.E. Mitzel (Ed.), Encyclopedia of educational research (Fifth edition). New York: The Free Press.

Goslin, D.A. (1967). Teachers and testing. New York: Russell Sage Foundation.

Hamilton, D., et.al. (Eds.) (1977). Beyond the numbers game. Berkeley, CA: McCutchan.

Hughes, W.H. (1934). Two educational philosophies: Are they compatible? Nation's Schools, 14(6), 25-28.

Jackson, P.W. (1968). Life in classrooms. London: Holt, Rinehart and Winston.

Jersild, A.T. (1929). Examinations as an aid to learning. Journal of Educational Psychology, 20, 602-609.

Lefever, D.W. (1950). The teacher's role in evaluating pupil achievement. Education, 71(4), 203-209.

Macken, I.N. (1929). Participation in testing programs by the classroom teacher. Educational Administration and Supervision, 15, 117-126.

Madaus, G.F. (1981). Reactions to the "Pittsburgh Papers." Phi Delta Kappan, 62(9), 634-636.

Main, J.E. (1930). History of the measurement movement. Educational Outlook, 4, 72-80.

Manning, W.H. (1969). The functions and uses of educational measurement. In P.H. Dubois (Chmn.), Proceedings of the 1969 Invitational Conference on Testing Problems. Princeton, NJ: Educational Testing Service.

Mayo, S.T. (1959). Testing and the use of test results. *Review of Educational Research*, 29(1), 5-14.

Messick, S. (1983). Abilities and knowledge in educational achievement testing: The assessment of dynamic cognitive structures. In S. Elliott and J. Mitchell, Jr. (Eds.), *Buros-Nebraska series on measurement and testing (Vol I)*; and B. Plake (Ed.), Social and technical issues in measurement and testing: Implications for test construction and usage. Hillsdale, NJ: Erlbaum.

Miller, D.F. (1963). State and national curriculums and testing programs: Friend or foe of expertness in the classroom. Journal of Secondary Education, 38, 41-47.

National Society for the Study of Education (1935). Thirty-fourth yearbook: Educational diagnosis. Bloomington, IN: Public School Publishing Company.

Peterson, J.J. (1983). Iowa Testing Program: The first half-century. Iowa City: University of Iowa Press.

Salmon-Cox, L. (1981). Teachers and standardized achievement tests: What's really happening. Phi Delta Kappan, 62(9), 631-634.

Scates, D.E. (1938). The improvement of classroom testing. Review of Educational Research, 8(5), 523-536.

Scates, D.E. (1947). Fifty years of objective measurement and research in education. Journal of Educational Research, 41, 241-264.

Snow, R.E. (1980). Aptitude and achievement. In W.B. Schrader (Ed.), Measuring achievement: Progress over a decade (Proceedings of the 1969 ETS Invitational Conference). San Francisco: Jossey-Bass.

Stinnett, T.M. (1969). Teacher certification. In R.L. Ebel (Ed.), Encyclopedia of educational research (4th ed.). New York: Macmillan.

Thelen, H.A. (1969). The evaluation of group instruction. In The sixty-eighth yearbook of the National Society for the Study of Education, Part II: The scientific movement in education. Chicago: University of Chicago Press.

Twenty-first Annual Conference on Educational Measurements (1934). Indiana University School of Education Bulletin, 11, 1-93.

Wood, B.D. (Ed.) (1933). Educational tests and their uses. Review of Educational Research, 3(1).

Weiss, J. (1980). Assessing nonconventional outcomes of schooling. In D.C. Berliner (Ed.), Review of research in education (No. 8). Washington, DC: American Educational Research Association.